Assessing California's Accountability System:

Successes, Challenges, and Opportunities for Improvement

American Institutes for Research®
Jennifer O'Day, Catherine Bitter

CPRE
Mike Kirst, Martin Carnoy

PACE
Elisabeth Woody, Melissa Buttles, Bruce Fuller, David Ruenzel

Policy Brief 04-2

POLICY BRIEF FEBRUARY 2004

Accountability for student performance is on the minds of everyone in American education—from policymakers to district administrators to principals. While the federal No Child Left Behind Act of 2001 (NCLB) has claimed center stage in the national accountability debate, California's own results-based accountability system was set in place several years prior to NCLB. In 1999, California legislators passed the Public Schools Accountability Act (PSAA), establishing specific performance targets for schools, a system of rewards and sanctions for meeting those targets, and assistance for low-performing schools. During the past two years, three independent studies—conducted by Policy Analysis for California Education (PACE), American Institutes for Research (AIR), and the Consortium for Policy Research in Education (CPRE)—have examined the accomplishments, shortcomings, and continuing challenges of PSAA. This brief outlines common findings and recommendations across these three independent studies.

Now is an opportune time to carefully assess the lessons of PSAA as California's new administration sets a path for improving instruction and student learning in light of recent developments:

Key Findings

These three studies reveal common patterns in teachers, principals, and district administrators experiences in implementing California's PSAA.

- Educators reported a renewed and unparalleled focus on improving student achievement and low-performing schools, also evidenced by rising test scores.
- The state's short-lived rewards and threats of sanctions were not an adequate motivating force for changing practice and increasing achievement. Educators responded more to public reporting of API scores and the scrutiny associated with the identification of schools as low-performing.
- Educators in the lowest-performing schools often reported insufficient resources or knowledge to significantly improve school practices and student learning. The state's efforts to increase the capacity of these schools were hampered by limited attention to implementation and follow-up.
- Effective use of data to inform practice was sporadic. Many educators lacked the skills, capacity, or time to analyze, understand, and respond to data on achievement gaps.
- District policy and context significantly influenced PSAA implementation, instructional improvement, and student achievement.

The Three Studies:

This brief draws on over 250 interviews with local educators conducted by the American Institutes for Research (AIR), Policy Analysis for California Education (PACE), and the Consortium for Policy Research in Education (CPRE). AIR's research documented, in particular, the impact of the Immediate Intervention/ Underperforming Schools Program (II/USP). AIR analyzed achievement data for all schools participating in Cohorts 1-3 of II/USP and comparison schools; administered surveys to teachers, principals, External Evaluators, and district staff; and conducted case studies in twenty-one urban, rural, and suburban schools that entered the II/USP between 1999 and 2001. PACE's Educator Responses to Accountability Project (ERAP) examined teachers' and administrators' experiences with the state's accountability system. The research centered on eight elementary schools within four districts—urban, suburban, and rural—across the state, interviewing educators about their understanding of the state system of accountability and its impact on classroom practices, and their efforts to address inequities in student achievement within the context of recent reforms. CPRE examined how nine high schools in six districts were responding to accountability pressures, including the expectation that students would eventually have to pass the California High School Exit Exam (CAHSEE) in order to receive a high school diploma.

- California's budget crisis has constrained the ability of the state and districts to provide assistance and incentives to schools. This will encourage policymakers to identify and build on the most successful aspects of current policies when prioritizing spending.
- Compliance with NCLB will require the state to resolve differences between its current processes for tracking school performance and aiding low-performing schools.
- The state's STAR testing and school assessment system must be reviewed and reauthorized this year, offering the legislature an opportunity to revisit the timing, components, and uses of the student testing system.
- Governor Schwarzenegger has proposed ways to simplify the state's school finance system, providing more flexibility to districts and school principals to allocate resources in ways that foster innovation and trust in local educators.

The purpose of this brief is to inform state and district-level policymakers by clarifying complex accountability issues and highlighting local educators' views that point to specific ways of improving California's accountability system.

Five key issues emerged during our three independent research studies in schools across California. This brief examines each issue, beginning with the relevant assumption underlying the accountability policy in that area —that is, how the policy was *intended* to work—and then discussing our common findings—that is, how the policy actually played out in practice. In this discussion, we consider both the successes and the continuing challenges of the system. We conclude by outlining four major implications for policymakers that result from these findings.

Attention to Student Achievement

■ ASSUMPTION: The state accountability system will focus the attention of teachers and administrators on improving student achievement, particularly in low-performing schools.

The development of California's accountability system was inspired in part by a sense that few people, including educators and parents, had a clear idea of just what students were learning and how well they were learning it. Schools and districts face multiple challenges in meeting the needs of students, particularly in low-performing schools that often have less experienced teachers, higher staff and faculty turnover, and parents struggling with economic and social hardships. One purpose of the state system of accountability has been to focus the attention of stakeholders on the primary goal of improving student achievement, particularly in low-performing schools. The assumption was two-fold: that standards would provide teachers with an understanding of what their students should learn over the course of an academic year; and

Components of California's System of Accountability

Academic Standards and Aligned Assessments

At the heart of California's accountability system are the academic content standards, specifying just what California students should learn, and annual student assessments that inform educators, policymakers, and parents as to how well students are meeting those standards. Initially, the system assessed students with an "off-the shelf" basic skills test, the Stanford 9; now, the core of the testing system is the California Standards Test, designed to align with the state's rigorous academic standards.

The Academic Performance Index (API)

California's accountability system provides educators, parents, and the public with information as to how well individual schools are performing compared to a state standard, in an effort to catalyze school improvement. The Academic Performance Index, or API, is a composite scale used to measure a school's performance and growth on the annual statewide assessment. Each school receives an API score, ranging from a low of 200 to a high of 1000, with an annual growth target of five percent of the distance between that school's current (baseline) API and the statewide interim goal of 800.

Growth targets are also set for numerically significant subgroups of students, including racial and ethnic groups, English learners, students with exceptional needs, and students who qualify for free or reduced price meals.

Rewards, Sanctions, and Assistance to Help Schools Improve

A central goal of the API is to provide educators with information on student achievement so that they can build on instructional strengths and address deficiencies. The API is also the mechanism by which schools are identified for rewards, sanctions and assistance.

During the time of our studies, schools that ranked in the bottom half of API scores statewide and failed to meet their annual growth targets could enter the Immediate Intervention/Underperforming Schools Program (II/USP). This program provided funds to support low-performing schools' efforts to improve, including one year for planning and two years for implementation. Schools participating in the II/USP would be subject to future sanctions (including takeover or even closure) should they not improve. The state postponed and/or limited sanctions to the assignment of state assistance and intervention teams. The II/USP has since been replaced by the High Priority Schools Grant Program (HPSGP), which targets assistance funds to the lowest-performing schools in the state.

Schools that met their achievement goals, on the other hand, were eligible for the Governor's Performance Award (GPA) program, which provided financial incentives as a reward for success. This program, and others like it, has been suspended as a result of the state's budget crisis.

that aggregate school API scores as well as individual student test scores would give administrators, teachers, and parents an understanding of how well students were mastering the standards and, by implication, how well schools were teaching them.

■ FINDING: The state accountability system has brought unparalleled attention to improving student achievement and low-performing schools.

Evidence from all three studies indicated that educators are aware of state standards and expectations for improvement, and had increased their focus on student achievement, especially in low-performing schools. District and school administrators acknowledged that the annual publication of the API scores has given them a measuring stick of student performance, and many teachers have a sense of how well their students are achieving and of areas that need improvement. Furthermore, district administrators were often focusing energy and

programs towards their lowest-performing schools. The increased attention to student achievement and standards likely accounts for the significant gains students and schools have made on the API since PSAA's implementation in 1999. This is particularly true at the elementary level, where schools have gained approximately 100 points since 1999, moving from a mean of 633 points in 1999 to 732 in 2003.

Within schools, teachers credited the state's academic standards with giving them a clear understanding of what they were to teach and of what students were supposed to learn. One elementary school teacher, for instance, said that the standards allowed her to "look at what I am supposed to teach, what students are going to be tested on, and know what is coming up." Teachers interviewed in each of these studies echoed this statement—they believed that the standards helped them map out measurable goals, making them aware of key concepts that they could then incorporate into lesson plans. At the high school level, academic departments expressed appreciation for how the standards helped bring focus to their meetings and instructional efforts.

Teachers often commented that the explicit nature of the standards played a direct role in motivating them to expect more from their students. One teacher, for instance, exposed his students to more difficult math concepts while a bilingual teacher became more aware of the need to motivate his students to read in English. And because the standards stipulated the same set of expectations for all students, they were expected to be a first step in helping teachers reduce disparities among high-poverty and minority students.

However, teachers also voiced numerous concerns over the standards. Some teachers, for example, felt that the standards were too extensive or developmentally inappropriate. The large numbers of topics, especially at the elementary school level, caused teachers in the PACE study to complain about having to adopt a "breadth over depth" approach to teaching, moving through the curriculum before students had a chance to fully master concepts. And teachers reported that the concepts themselves were often calibrated too high; appropriate perhaps for children who came to school "prepared to learn" but often overwhelming for others who fell far behind. One teacher, for instance, reported that a fourth grade standard required students to master algebra with expressions and variables. With even her best students struggling with these advanced concepts, she felt conflicted over the appropriateness of the required material.

The policy's success in focusing attention on students' achievement in math and language arts may also have had unintended repercussions. Educators across the three studies raised concerns that the emphasis on basic reading and mathematics skills combined with prescriptive curriculum packages limited their ability to tend to other subjects such as social studies, science, and the arts. Some teachers also felt that the strong focus on standards and test scores constrained their pedagogical choices and forced them to spend too much time "teaching to the test."

Our interviews with local educators also found mixed levels of understanding of the state's accountability system among teachers. Teachers varied in their understanding of and ability to apply API data to their work. For instance, while some teachers used their API scores to inform their practices, other teachers simply felt APIs scores were not particularly useful. One teacher in the PACE study noted, "It's [API score] not something I put a lot of time into. I could get it for you and tell you what our score is, but it's not something that I spend a lot of time analyzing." Furthermore, teachers were not always familiar with their state and similar schools rank, or their subgroup scores, and in several cases were unaware that disaggregated subgroup data existed at all.

Adding to the confusion were the new requirements set forth by the federal NCLB legislation, which were just being introduced into schools during the time of our research. The overlapping of the two systems, which was less than seamless, often added to educators' confusion regarding accountability.

Rewards and Sanctions as Motivators

■ ASSUMPTION: Rewards and sanctions, tied to student test scores, as well as public scrutiny, will motivate educators to improve outcomes.

The school accountability system in California, like that in many other states, operates on the assumption that educators, like employees in other public and private institutions, will respond to a coherent set of rewards and sanctions. The theory of results-based school accountability suggests that teachers and administrators, as well as students, need *extrinsic* motivation to continually strive to meet API targets.

California policymakers assumed that improved school scores, reported by the media, would "reward" educators with enhanced recognition and public esteem; poor scores, on the other hand, would motivate educators to prevent the label of teaching in low-achieving schools. Additional motivation would come from the promise of rewards and the threat of severe sanctions. Educators would work hard, for instance, to receive the cash rewards the state offered for high achievement gains. They would also work to avoid sanctions, such as possible school takeovers and/or closures.

■ FINDINGS: School personnel did respond to the identification of schools as low-performing and to the publication of API scores and ranks. However, these studies found little evidence that either the monetary rewards or the threatened severe sanctions tied to student test scores provided effective incentives to leaders or schools.

Awards

Although teachers and administrators in schools that received awards such as the Governor's Performance Award (GPA) were pleased to receive additional funds, the rewards did not appear to be a strong motivating factor among educators. For example, high school teachers in CPRE's study did not mention rewards at all; they were concerned with other pursuits, such as preparing students to pass the new California High School Exit Exam. And while other teachers usually had at least a hazy conception of the programs, they were dismissive of their significance.

In many cases, teachers did not expect that awards would actually be given if they improved student outcomes. For example, when asked how likely it would be for their school to receive a financial reward from the state if they met their API growth targets, only 27 percent of teachers surveyed in AIR's study (including II/USP, comparison, and higher achieving schools) reported that it was likely to happen or definitely would happen. Similarly, only 24 percent of principals reported that it was likely to happen or definitely would happen. This lack of expectation likely came in part from the delayed distribution of funds and the short-lived nature of the programs.

Altogether, four different rewards programs operated at one time, all of which are now suspended or defunct. The most prominent rewards program, the GPA, distributed cash monies to schools that demonstrated high achievement only in 2001 and 2002; the program was then stopped due to the state's budget crisis. As one teacher from AIR's study said, "most of that money never came to anyone anyway, so it was all a smokescreen anyway, and we knew that coming into it."

Other teachers, even at schools that did receive cash rewards, were skeptical of their value. "And the other thing that is really annoying is the whole paying teachers for getting test scores up," one elementary teacher said. "It's just going to lead to cheating, or teachers are going to do what they can to get that money. . . ." Another teacher put the attitude towards awards in perspective when he said, "I am not in education to be paid more because my kids all scored well on the test. I am in education because I want to help these kids grow up to be better people and grow up to learn."

Sanctions

The threat of severe sanctions also proved to carry little saliency among educators. Fewer than half of the teachers in II/USP schools surveyed in AIR's study believed that the most severe sanctions would occur, even if their school failed to make the requisite progress. They saw the state's educational policies as being in perpetual flux. One principal in AIR's study explained teachers' lack of concern regarding a potential state takeover: "most of them did not take

that threat very seriously, because they said, 'they'll never do it, they never do it." Instead, many believed that less severe sanctions, such as a public hearing or state assistance team, would be implemented. Educators in the other studies echoed similar expectations.

These findings do not imply that rewards and sanctions are inherently ineffective. Because the rewards and sanctions were not implemented consistently, and because educators doubted they would actually occur, it is difficult to tell whether their general ineffectiveness lies in the theory of accountability itself or in the uneven execution of that theory. Much more effective in terms of motivating California's teachers was the sheer weight of public scrutiny. Teachers wanted to feel proud of their schools, their students, and themselves; the recognition that attended the publication of their school's API score was truly a motivating force for many.

School Capacity

■ ASSUMPTION: Many schools have sufficient capacity to improve student achievement. The additional resources and assistance provided through the II/USP would help the lower-capacity schools improve student outcomes through additional funds, better planning, and external assistance.

California's school accountability system assumes that many schools have sufficient resources (e.g., funding, qualified staff, school leadership) to improve student achievement if attention and incentives are directed toward this goal. For those schools without sufficient resources, policymakers believed that providing additional funds and external assistance would increase capacity sufficiently to improve student outcomes. The II/USP, for example, provided low-performing schools with an External Evaluator, an "expert" funded by the state who would help educators develop a comprehensive Action Plan for school improvement (as measured by student gains on the state's achievement test). Such schools also received additional funds with which to implement the improvement plan.

California policymakers also assumed that educators would have the necessary time and skills to analyze and interpret the data from the annual assessment. Educators would then note inequitable trends in student achievement and address them with data-driven instruction.

■ FINDINGS: The accountability system does not adequately address the limited capacity of the lowest-performing schools to improve student learning or to use data effectively, even with the additional assistance provided from the state.

The state's accountability system assumes that administrators, teachers, and students within individual schools have the capacity to change behaviors in ways that support learning. But while the three studies found that some schools showed an admirable ability to meet the needs of their students, others—particularly the lowest-performing schools—simply lacked sufficient capacity, including a strong professional community, instructional leadership, and material resources, to make meaningful and sustainable improvements in student achievement.

Schools serving California's most disadvantaged students, for instance, were often hampered by intensive teacher turnover, making it difficult for them to put long-term improvement plans into action. Often, they lacked the "basics" of instructional materials, safe and comfortable facilities, and time for professional development. Although the II/USP provided funds to many of these schools for improvement, the funds were often not enough to make significant changes and/or not provided in a timely manner.

The External Evaluator provision of II/USP was expected to build capacity at schools by helping schools to identify areas of need during the initial planning year and to develop a coherent plan for improvement. AIR's evaluation found that although the planning activities typically conformed to the requirements of the legislation, the quality and depth of the planning experiences varied greatly, as did the quality and capacity of the External Evaluators and their organizations.

AIR's research found that the II/USP's effect on student achievement was hampered by the frequent disconnect between planning that took place with the External Evaluator

Accountability and the California High School Exit Exam

In most respects, high school educators responded to accountability pressures in much the same way as did their elementary and middle school counterparts. Administrators and teachers were both extremely aware of the state's academic standards and the API system, though administrators often attributed greater importance than teachers to increasing API scores. The pressures of the API growth targets were greatest on low-performing schools; schools in the middle of the API range seemed less concerned and were generally not making a concerted, school-wide effort to improve scores.

One factor did, however, differentiate the high school response to accountability—the high stakes attached to the California High School Exit Exam (CAHSEE). CPRE researchers found that this standards-based exam, which students originally had to pass beginning in 2004—that date has now been pushed back to 2006—in order to receive a high school diploma, was inspiring a number of important educational initiatives. This was especially true in schools where a large number of students were in danger of failing the exam. Districts generally encouraged and even mandated some of these initiatives. Some schools, for instance, were offering after-school tutoring and test-preparation classes; others required a class period in which students were taught skills and content reflected in the exam. A number of schools, realizing that their students lacked the reading skills needed to pass the exam, implemented basic reading programs.

In short, many student achievement initiatives were undertaken in response to the high school exit exam—initiatives that may or may not be on shaky ground now that the full implementation of the exam has been postponed until 2006. We cannot predict how California educators will respond to this postponement. The hope is that educators will continue to focus on improving achievement to meet the demands of the exam. But the postponement could induce skepticism as to the validity of the 2006 date, causing educators to lose that focus.

The exam was postponed by the state board of education in part because an independent evaluator, appointed by the state as required by law, indicated that there were insufficient opportunities in many schools to learn the material on the exam. This finding dovetails with the contention, explored in the capacity section of this brief, that many schools lack the resources and expertise to significantly improve student achievement. While the accountability system is centered on getting students to meet academic standards, there are few capacity-building opportunities built into the system.

in the first year and subsequent implementation of the plans. The policy did not require consistent monitoring, assistance, or follow-through after the planning year, making it difficult for schools to follow through with a coherent instructional plan, an important contributor to long-term school improvement. Both CPRE and AIR's studies showed that when an external assistance provider was hired to assist with the implementation years, the school demonstrated greater success in implementing its Action Plan.

The need for higher capacity in many schools was also made evident through teachers' attitudes about what they could accomplish in the classroom. Many teachers interviewed felt that they were dealing with factors beyond their control. One teacher in the PACE study explained her frustration with the state's expectation that schools were capable of closing achievement gaps, noting, "it is ridiculous because it is not like we are given tools with which to help correct the problem." Unfortunately, it was not uncommon for teachers to excuse low achievement among Latinos and African Americans by

pointing to a lack of parent support or community culture that did not value education. A focus group discussion among teachers and administrators at one school in PACE's study, for instance, revealed the perception that African American students at that school were simply more interested in basketball than academics.

One point of the accountability system is to change these attitudes, creating a situation in which schools do not blame students and their background for continued low achievement, and instead take the responsibility to improve the learning of all students. To do so, however, educators need to understand the data about the gaps that exist, have the knowledge and skills to devise solutions, and have the resources to put those solutions into practice. If those conditions aren't present, it will be difficult to foster improvement, and continued low performance of students may "confirm" any pre-existing ideas about their inability to learn or perform to high standards.

These studies found that teachers and other educators often had neither the capacity to understand and use the disaggregated data nor the capacity to create feasible solutions to identified problems. As one teacher told us, "We do a lot with what we have. And to me the natural consequence is that they're going to improve and you're going to reach a certain point. And then, based on your demographics, you're not going to improve anymore." Even when sufficient human capacity was present, the requisite resources to make those solutions a reality were often not forthcoming or were insufficient for the depth and breadth of the problem. This could lead to a sense of frustration, even futility, as educators come to find the expectations of the accountability system unrealistic.

Using Achievement Data to Improve Instruction

■ ASSUMPTION: The accountability system will provide standardized test data that will be used to show overall school and student performance, highlight achievement gaps, and help teachers improve student performance. Teachers and schools will be able to use these data to make informed instructional decisions.

The system assumes that data from statewide standardized testing will be valuable in assessing performance in a number of ways:

- Educators will know how students at their school are achieving relative to students at schools with similar demographics and resources;
- They will know how their school is performing relative to an absolute standard, namely an API score of 800;
- They will know how their school's performance has changed over time;
- And they will know, with disaggregated data, how numerically significant subgroups of students at their school are achieving.

Data will be important in helping administrators and teachers assess their school's performance and work toward improving student achievement. Policymakers expected the disaggregated subgroup data to contribute to the improvement of achievement for students that have long been lagging behind. The assumption is that educators can drive that improvement by using the data to illuminate discrepancies and to construct solutions to close gaps in achievement.

■ FINDINGS: Standardized test data have been effective in showing overall academic performance, but have been less than fully effective in helping educators make data-based instructional decisions.

While the API data did provide a broad picture of student achievement, the data were not necessarily useful in helping teachers develop specific instructional strategies for improvement. Local educators often lacked the knowledge base and data management tools to make full use of the data. In addition, knowledge of student outcomes does not guarantee that educators will know what changes need to be made in instruction to reach those outcomes. As one teacher in the PACE study explained, "There's a difference between taking information and using it to try to determine why it's so and if something can be done."

One problem that emerged during PACE's research was that not all teachers had good knowledge of their school's subgroup data, and hence could not target the particular needs of these groups. In some schools teachers did not even know that data existed in a disaggregated form;

others knew of its existence, but not of the details for their school. Most teachers were familiar with subgroup data only in the context of whether the school had met its subgroup targets that year. They might know, for example, that their Latino population had gone up by thirty points, but that their Asian population had missed their target by one point. But it was less common to hear teachers involved in an ongoing analysis of disaggregated data to inform instruction.

Teachers also claimed that the fact that test scores were not released until the end of summer subtracted from their usefulness. By August, the students they taught were moving on to new teachers. Some evidence suggested that more regularly administered assessments, from which teachers received steady feedback, were valuable in terms of informing and modifying instruction. For instance, teachers in one school in the PACE study and several schools in the AIR study used a practical assessment program that evaluated student achievement in smaller, more useful increments. As one teacher in the PACE study noted: "We're using assessments in smaller increments to see student growth instead of, 'Oh the next grade level, what do we do with them?' And we're really trying to assess students and change our instruction while they're still in our classroom."

By receiving immediate scores and feedback for each student, teachers had multiple opportunities to refocus instruction and re-emphasize key standards, to which the program was aligned. Similarly, a high school district in the CPRE study tested students in all its schools at the beginning and end of each school year in reading and math, which provided a baseline to evaluate incoming students and the ability to assess value added over the course of the year.

Finally, many of the schools lacked a culture that supported and encouraged the use of data. Administrators, as well as teachers, generally had little educational background and practical experience in breaking down and analyzing data. At some schools, for instance, faculty exposure to subgroup data was limited to a quick examination of scores posted by the principal. Schools that studied data seriously, on the other hand, typically had a strong principal who helped the faculty draw valuable information from it.

API data, when teachers did study it, often only confirmed what they already knew—namely that certain subgroups in their school were lagging behind. The end result was simply another exposure of pre-existing school failures or successes. As one teacher told us, "I suppose if we could get the answers [as to how to close achievement gaps] that'd be great, but what if we couldn't? What if it really meant, 'sorry, they don't do as well'?"

It is important to note that there were a few occasions when schools did use data effectively. At one school in the PACE study, for instance, the principal used data to argue the need for after-school or Saturday programs for low-performing students: "That [Saturday program] would never have passed through the PTA board or the teachers if we didn't have the data to show the specific target group." Another school's teachers used data from frequently administered "low stakes" tests to note achievement gaps and regularly adjust their instruction to address them.

The Role of the District

■ ASSUMPTION: Schools are the primary unit responsible for student achievement, and therefore the primary unit for change.

As we have emphasized throughout this brief, California's accountability system assumes that the individual school has the chief responsibility for improving student achievement and should face consequences for failing to do so. The district, on the other hand, plays a supporting role. While the district can offer guidance and assistance, it is not responsible—and not held accountable—for school and student performance.

■ FINDINGS: Districts often play a major role in determining how effective schools can be in improving student achievement.

We found that the school was far from the independent unit the developers of the accountability system conceived it to be. Districts influenced how schools responded to accountability in many ways. They frequently determined, for instance, which

schools participated in the II/USP, and sometimes played an active role in selecting the External Evaluator. Some districts set up required supports for the implementation of the Action Plan, including external assistance, professional development, and mentoring.

Districts had strong influences on nearly every school in the three studies—not just those entering the intervention program. AIR's analysis of achievement data in the four districts with the largest number of II/USP schools—Los Angeles, San Diego, San Francisco, and Oakland Unified—revealed substantial contributions (positive and negative) of district membership on student achievement growth in both II/USP and comparison schools. Information from all three studies reveals that this influence came in large part through instructional policies, which districts implemented for all of their schools or for those designated as underperforming.

Some districts, for instance, took a very active role in curriculum and instruction; in particular, they often mandated the use of a common literacy program that would, they hoped, raise achievement scores and bring coherence to schools' instructional approaches. And indeed, AIR's study found that reading scores generally did rise, sometimes significantly, in schools that implemented more specific, coherent approaches.

On the other hand, teachers in some schools in both the PACE and AIR studies reported that district policies occasionally hampered their instructional efforts. They felt, for one thing, that the curriculum districts sometimes mandated reduced their professional autonomy and ability to meet student needs. One teacher commented that a district's pacing plan required teachers to teach "Open Court for three hours and you have to be on a certain story, and you have to teach one math lesson a day. And if your kids don't get it you are supposed to move on when you can't, and then you get so behind that they have to take a test that they are not ready for." In addition, teachers complained about the district's Open Court "police"—district officials who enforced full compliance with the reading program even when such enforcement, teachers felt, interfered with best practice. Teachers in some districts were also concerned about additional district assessments of students, adding to an already heavy testing burden.

It is important to emphasize, however, that regardless of how involved districts were in school improvement efforts, or how effective or ineffective their improvement efforts might have been, they were not directly accountable for school successes or failures. At the time of our research, PSAA's accountability burden fell solely on schools.

Where Do We Go from Here? Implications for California Policymakers

What are the practical implications of our findings, as California policymakers consider adjustments to the state's system of accountability? It is our hope that state-and district-level policymakers can continue to work to improve public school accountability, through the following measures:

Ensure stability and consistency within the system of accountability

Since the introduction of PSAA in 1999, the state's accountability system has undergone numerous changes, creating confusion and a sense among many educators that "this too shall pass." For example, the discontinuation of performance rewards only reinforced the skepticism of educators who believe that what is decided upon one year may be undone the next.

Policymakers should now build confidence in the system by sustaining those aspects that are proving effective and ensuring that any new initiatives are well thought-out so that they can endure over time. The state curriculum standards and aligned tests, for example, should remain in place to ensure teacher "buy-in." Likewise, efforts should be made to restore a system of rewards. While financial incentives may no longer be realistic, the state could offer incentives through regulatory relief for schools that show sustained growth over two or

more years, as recently proposed by the Superintendent of Public Instruction.

Efforts should also be taken to dispel any confusion about the intersection of the state's accountability system and more recent federal NCLB mandates. Educators and parents need a clear understanding of the different means by which the two systems measure school success and failure, and the implications of both systems for subsequent school sanctions. The state's system measures a school's growth in achievement by points on the API scale, whereas the federal AYP (Adequate Yearly Progress) system measures growth by percentage. It is entirely possible for a school to meet its target for growth based on the API, but not the AYP percentage, or vice versa. California's Department of Education and the Governor's office might consider negotiating with federal policymakers for a more simplified system that would build upon existing state measures and reduce additional confusion.

Strengthen policies to build the capacity of the neediest and lowest-performing schools

Target funds to the neediest schools

Our research suggests that providing small sums of additional money to a large number of underachieving schools was not the most effective use of resources, as such schools remained limited in their capacity to improve. NCLB presents an additional challenge, as more schools are likely to be designated as low-performing. It will become increasingly important for the state to direct intensive assistance efforts towards the lowest-performing schools.

Narrowing the group of targeted schools to those most in need is particularly important during this time of budgetary crisis. The state has recently taken this approach through the High Priority Schools Grant Program. The state may also rely on federal funding sources, as NCLB provides a potentially useful resource through the allocation of additional federal Title I funds and other funding for school improvement.

Provide *sustained* and effective support

However, our research suggests that a single infusion of money alone is not enough to ensure sustained improvement. Instead, the state should provide ongoing assistance to allow schools to develop and implement a coherent instructional program, including follow-up and resources beyond the initial planning phase. In addition, district policies should support, not hinder, school improvement efforts. District policies around teacher assignment or allocation of resources, for example, may influence a school's capacity to improve.

Meaningful financial assistance and support can build the capacity of low-performing schools through such initiatives as professional development and more coherent instructional programs. Professional development can be highly effective when teachers are consulted about its timing and topics, and when training supports a coherent, standards-based instructional program, not merely the implementation of an isolated curriculum package. State and district leaders can work together to establish guidelines for effective professional development, while ensuring schools the flexibility to address the needs of their specific communities.

Combine flexible funding with capacity-building

The state is now considering the use of a weighted-student approach to school funding, as well as a consolidation of categorical funds to simplify funding streams. Any adjustments to the current school funding system should provide schools with the flexibility to meet the needs of their specific populations, while also recognizing that schools may not always have the capacity to determine the best possible use of funds. School site planning under II/USP at times led to fragmented spending of discretionary funds. In conjunction with efforts to simplify funding, the state should issue guidelines for school spending, highlighting the need to build school capacity for improvement.

Finally, it is important to recognize that an external system of assessments cannot, in and of itself, drive educators in the lowest-performing schools to improve student achievement. Moreover, it cannot compensate for a lack of resources. If educators in the lowest-performing schools are to be consistently successful in

helping students to reach achievement goals, the state must ensure that they have the capacity to do so with a baseline of resources and support. The state has taken a step in this direction with efforts to develop opportunity to learn indicators to ensure that schools have the necessary resources to meet the demands of the state's system of accountability.

Facilitate more effective use of data and assessments

Data from the state's standardized tests have been useful in providing the public and educators with a picture of school performance and student achievement. However, these data are based on an assessment administered once a year and hence represent a snapshot in time. More frequent diagnostic assessments can allow teachers to identify learning problems and adjust instruction accordingly. The state Board of Education could encourage districts to develop and support such assessments, generating user-friendly data for teachers.

Our research also demonstrated that many educators lack opportunities to review student test data. Moreover, they lack a clear understanding of how to interpret and analyze the data, and how to apply what they learn to improve classroom practices. There is a need for schools to develop a culture that supports the use of data to improve instruction. It is essential that principals be trained in this area and that teachers receive professional development not only to provide data analysis skills, but to demonstrate practical classroom applications of student achievement data.

Enhance the role of districts within a system of accountability

PSAA focused accountability mandates almost entirely at schools. In response to NCLB requirements, the state must now design a system to hold districts accountable as well. These three studies have shown that districts are already engaged in the processes of school accountability. Any system of district accountability should recognize district involvement, and encourage districts to support, not hinder, school efforts to improve.

Districts are, or should be, in a position to offer valuable support. They can help schools use assessments and data more effectively, target resources to where they are most needed, and offer professional development to foster improved instruction. If districts are to better support schools, the state, in turn, must better support districts. The state should do what it can to streamline funding and provide guidelines for effective resource allocation.

Five years after the passing of PSAA, California's accountability system faces a time of both reflection and renewal. California's schools have made significant gains in student test scores since the implementation of PSAA, particularly at the elementary level. Questions remain as to whether the system can support sustained and additional growth. The system has also undergone significant changes over the last few years, including a shift to a standards-aligned test, and the elimination of outcomes-based financial rewards for educators. Additional changes may be inevitable, as state policymakers face decisions regarding school funding and the reauthorization of the STAR testing system in the coming year.

Through these three independent studies, teachers, principals, and district leaders throughout the state have provided insight into those aspects of the state system of accountability that best support school improvement efforts. It is our hope that any adjustments to California's accountability system build on lessons learned from these studies and continue to support educators' efforts to improve student learning.

American Institutes for Research®

CPRE

Relevant Publications

- Woody, E.L., et al. ***"Voices from the Field: Educators Respond to Accountability."*** February 2004
- Novack, J.R., Fuller, B. ***"Penalizing Diverse Schools? Similar test scores, but different students, bring federal sanctions."*** December 2003

To obtain copies, or download reports, please visit the PACE website: http://pace.berkeley.edu.

- O'Day, J., Bitter, C. ***"Evaluation Study of the Immediate Intervention/Underperforming Schools Program and the High Achieving/Improving Schools Program of the Public Schools Accountability Act of 1999."*** June 2003

To obtain copies, or download reports, please visit the AIR website: www.air.org

- For further information on CPRE's publications, please visit the CPRE website: www.cpre.org

Acknowledgements:

This project was generously funded by the Noyce Foundation and the William and Flora Hewlett Foundation. Special thanks to Jeannie Oakes for her valuable insight and contributions to this effort.